JET PROPULSION LABORATORY

NOTIFICATION OF CLEARANCE

10/30/02

TO: H. Zima

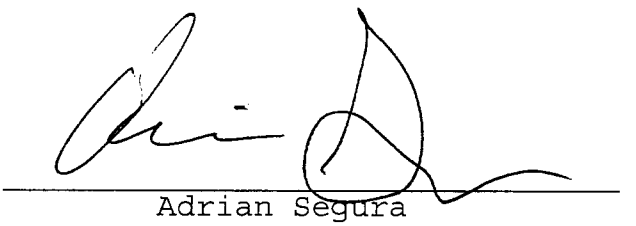
FROM: Logistics and Technical Information Division

SUBJECT: Notification of Clearance - CL#02-2786

The following title has been cleared by the Document Review Services, Section 274, for public release, presentation, and/or printing in the open literature:

The Cascade Programming and Execution Model: A First Approach

This clearance is issued for the full paper and is valid for U.S. and foreign release.

Clearance issued by 

Adrian Segura
Document Review Services
Section 644

(Over)

34424

JPL AUTHORIZATION FOR THE EXTERNAL RELEASE OF INFORMATION

Submit web-site URL or two copies of document with this form to Document Review, 111-120, or email them to docrev@jpl.nasa.gov.

CL No 02-2786 (for DRS use only)

LEAD JPL AUTHOR		MAIL STOP	EXTENSION
Zima	Hans P	126-201	8183548980

The Document Review approval process applies to all JPL information intended for unrestricted external release via print or electronic media. See explanations on page 3 of this form and the Distribute Knowledge documents available through http://dmie.

☒ Original ☐ Modified

I. DOCUMENT AND PROJECT IDENTIFICATION - To be completed by Author/Originator

☐ ABSTRACT (for publication) ☐ WEB SITE ☒ ORAL PRESENTATION

☐ FULL PAPER (including poster, video, CD-ROM) ☐ OTHER _____ ☐ Abstract ☐ Full Text

TITLE	OTHER AUTHORS	
The Cascade Programming and Execution Model: A First Approach	David Callahan, Cray, Inc.	☐ Premeeting publication ☐ Publication on meeting day ☐ Postmeeting publication ☐ Poster session ☐ Handouts

KEY WORDS FOR INDEXING *(Separate terms with commas)*

THIS WORK:
☐ Covers new technology not previously reported
☐ Covers work previously reported in New Technology Report (NTR) No. _____
☐ Provides more information for earlier NTR No(s). _____
☒ Contains no new technology per author 10/24/02

LEAD JPL AUTHOR'S SIGNATURE — DATE

SECTION OR PROJECT LEVEL APPROVAL - I attest to the technical accuracy of this document/web site. — DATE

Terry A. Bergman — 10/23/02

ORIGINATING ORGANIZATION *(Section, Project, or Element Number)*: 3660

PERFORMING ORGANIZATION *(If different)*:

ACCOUNT CODE OR TASK ORDER *(For tracking purposes only)*	DOCUMENT NUMBER(S), RELEASE DATE(S)	DATE RECEIVED	DATE DUE
		10/24/02	

For presentations, documents, or other scientific/technical information to be externally published (including via electronic media), enter information--such as name, place, and date of conference; periodical or journal name; or book title and publisher -- in the area below.

Web Site: Preclearance URL *(JPL internal)* _____

Postclearance URL *(external)* _____

☐ Brochure/Newsletter ☐ JPL Publication Section 274 Editor *(If applicable)* _____

☐ Journal Name _____

☒ Meeting Title Southern California Workshop on Parallel and Distributed Processing and Architecture

Meeting Date 10/28/2002 Location Santa Barbara, California

Sponsoring Society _____

☐ Book/Book Chapter ☐ Assigned JPL Task ☐ Private Venture Publisher _____

If your document will not be part of a journal, meeting, or book publication (including a web-based publication), can we post the cleared, final version on the JPL worldwide Technical Report Server (TRS) and send it to the NASA Center for Aerospace Information (CASI)? ☐ Yes ☐ No

(For more information on TRS/CASI, see http://techreports.jpl.nasa.gov and http://www.sti.nasa.gov.)

If your document will be published, the published version will be posted on the TRS and sent to CASI.

II. NATIONAL SECURITY CLASSIFICATION

CHECK ONE *(One of the five boxes denoting Security Classification must be checked.)*

☐ SECRET ☐ SECRET RD ☐ CONFIDENTIAL ☐ CONFIDENTIAL RD ☒ UNCLASSIFIED

III. AVAILABILITY CATEGORY - To be completed by Document Review

NASA EXPORT-CONTROLLED PROGRAM STI

☐ International Traffic in Arms Regulations (ITAR)
☐ Export Administration Regulations (EAR)

Export-Controlled Document -- U.S. Munitions List (USML Category) _____ or
Export Control Classification Number (ECCN) _____ from the
Commerce Control List (CCL) _____

CONFIDENTIAL COMMERCIAL STI
(Check appropriate box below and indicate the distribution limitation if applicable.)

☐ TRADE SECRET ☐ Limited until *(date)* _____
☐ SBIR ☐ Limited until *(date)* _____
☐ COPYRIGHTED ☐ Limited until *(date)* _____
☐ COPYRIGHT TRANSFERRED TO: ☐ Publicly available *(but subject to copying restrictions)*

ADDITIONAL INFORMATION
(Check appropriate distribution limitation below and/or limited until (date), if applicable.)

☐ U.S. Government agencies and U.S. Government agency contractors only
☐ NASA contractors and U.S. Government only ☐ U.S. Government agencies only
☐ NASA personnel and NASA contractors only
☐ Available only with the approval of issuing office ☐ NASA personnel only

☒ PUBLICLY AVAILABLE STI

Publicly available means it is unlimited and unclassified, is not export-controlled, does not contain confidential commercial data, and has cleared any applicable patent application.

IV. DOCUMENT DISCLOSING AN INVENTION (For SIAMO Use Only) ROUTED ON

☐ If STI discloses an invention Check box and send to SIAMO | COMMENTS

THIS DOCUMENT MAY BE RELEASED ON *(date)* | STRATEGIC INTELLECTUAL ASSETS MANAGEMENT OFFICE (SIAMO) SIGNATURE | DATE

IV. BLANKET AVAILABILITY AUTHORIZATION (Optional)

☐ All documents issued under the following contract/grant/project number may be processed as checked in Sections II and III.
This blanket availability authorization is granted on *(date)*_____ Check one: ☐ Contract ☐ Grant ☐ Project Number _____

The blanket release authorization granted on *(date)*_____
☐ is RESCINDED – Future documents must have individual availability authorizations.
☐ is MODIFIED – Limitations for all documents processed in the STI system under the blanket release should be changed to conform to blocks as checked in Sections II and III.

SIGNATURE | MAIL STOP | DATE

V. PROJECT OFFICER/TECHNICAL MONITOR/DIVISION CHIEF REVIEW OF I THROUGH V

☐ Approval for distribution as marked above ☐ Not appoved

NAME OF PROJECT OFFICER OR TECH. MONITOR | MAIL STOP | SIGNATURE | DATE

VII. EXPORT CONTROL REVIEW/CONFIRMATION ROUTED ON

☐ Public release is approved ☐ Public release not approved due to export control ☐ Export-controlled limitation is not applicable
☐ Export-controlled limitation is approved ☐ Export-controlled limitation (ITAR/EAR marked in Section III is assigned to this document)

USML CATEGORY NUMBER (ITAR)	CCL NUMBER, ECCN NUMBER (EAR)	JPL EXPORT CONTROL ADMIN. REPRESENTATIVE SIGNATURE	DATE

COMMENTS

VIII. OTHER APPROVALS ROUTED ON

☐ LAUNCH APPROVAL
☐ OFFICE OF COMMUNICATIONS AND EDUCATION
☐ GENERAL COUNSEL
  ☐ Budgetary/Cost Data
  ☐ Vendor Data
  ☐ Copyrights
  ☐ Other _____
☐ OTHER _____

COMMENTS

SIGNATURE | DATE

IX. FINAL VERIFICATION, APPROVAL, AND DISPOSTIION BY DOCUMENT REVIEW

I have determined that this publication:

☐ DOES contain ITAR/export-controlled, confidential commercial information, and/or discloses an invention and the appropriate limitation is checked in Sections III and/or IV.

☒ Does NOT contain ITAR/export-controlled, confidential commercial information, nor does it disclose an invention and may be released as indicated above.

USML CATEGORY NUMBER (ITAR) 120.11 | CCL NUMBER, ECCN NUMBER (EAR) _____

☒ Public release is approved for U.S. and foreign distribution ☐ Public release is not approved

COMMENTS

SIGNATURE [signature] | MAIL STOP 111-120C | DATE 10/30/02

☐ Obtained published version Date ☐ Obtained final JPL version Date

See page 3 for instructions for completing this form.

To: docrev@jpl.nasa.gov
Subject: Authorization for External Release

Hi!

Attached are the document and form 1330-S for your review and release for Hans Zima's presentation at UC Santa Barbara on next Monday, Oct. 28.

Your approval is appreciated!

~~~~~~

Winnie Wang <winnie.p.wang@jpl.nasa.gov>
Engineering and Communications Infrastructure/Sec. 366
Phone: 818/354-9856 ~*~ Fax: 818/393-0479 ~*~ MS: 126-256

~~~~~~

 socal.02a.ppt

 Cascade-Hans.tif

The Cascade Programming and Execution Model: A First Approach

David Callahan
Cray Inc., Seattle, Washington
and
Hans P. Zima
NASA Jet Propulsion Laboratory, Pasadena, California

Southern California Workshop on Parallel and Distributed Processing and Architecture
Santa Barbara, California
October 28, 2002

Outline

- 1 The DARPA HPCS Program
- 2 The Cascade Project
- 3 Cascade Hardware Architecture
- 4 Basic Programming Model
- 5 Extended Programming Model
- 6 Irregular and Dynamic Applications
- 7 Research Issues
- 8 Conclusion

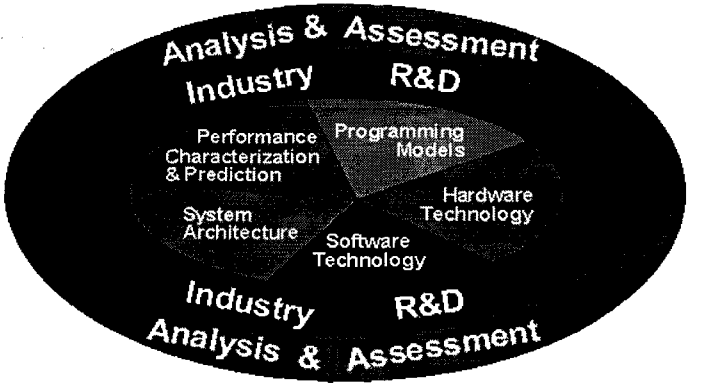
High Productivity Computing Systems

Goals:

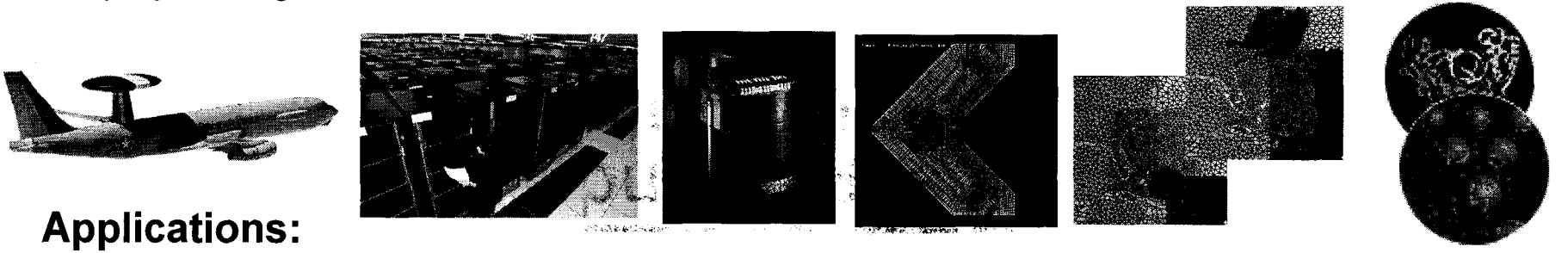
- **Provide a new generation of economically viable high productivity computing systems for the national security and industrial user community (2007 – 2010)**

Impact:

- **Performance** (efficiency): critical national security applications by a factor of 10X to 40X
- **Productivity** (time-to-solution)
- **Portability** (transparency): insulate research and operational application software from system
- **Robustness** (reliability): apply all known techniques to **protect against outside attacks**, hardware faults, & programming errors



HPCS Program Focus Areas



Applications:

- Intelligence/surveillance, reconnaissance, cryptanalysis, weapons analysis, airborne contaminant modeling and biotechnology

Fill the Critical Technology and Capability Gap
Today (late 80's HPC technology)….*to*…..Future (Quantum/Bio Computing)

The Cascade Project

- 1-year Concept Study, July 2002-June 2003
- Led by Cray Inc. (Burton Smith)

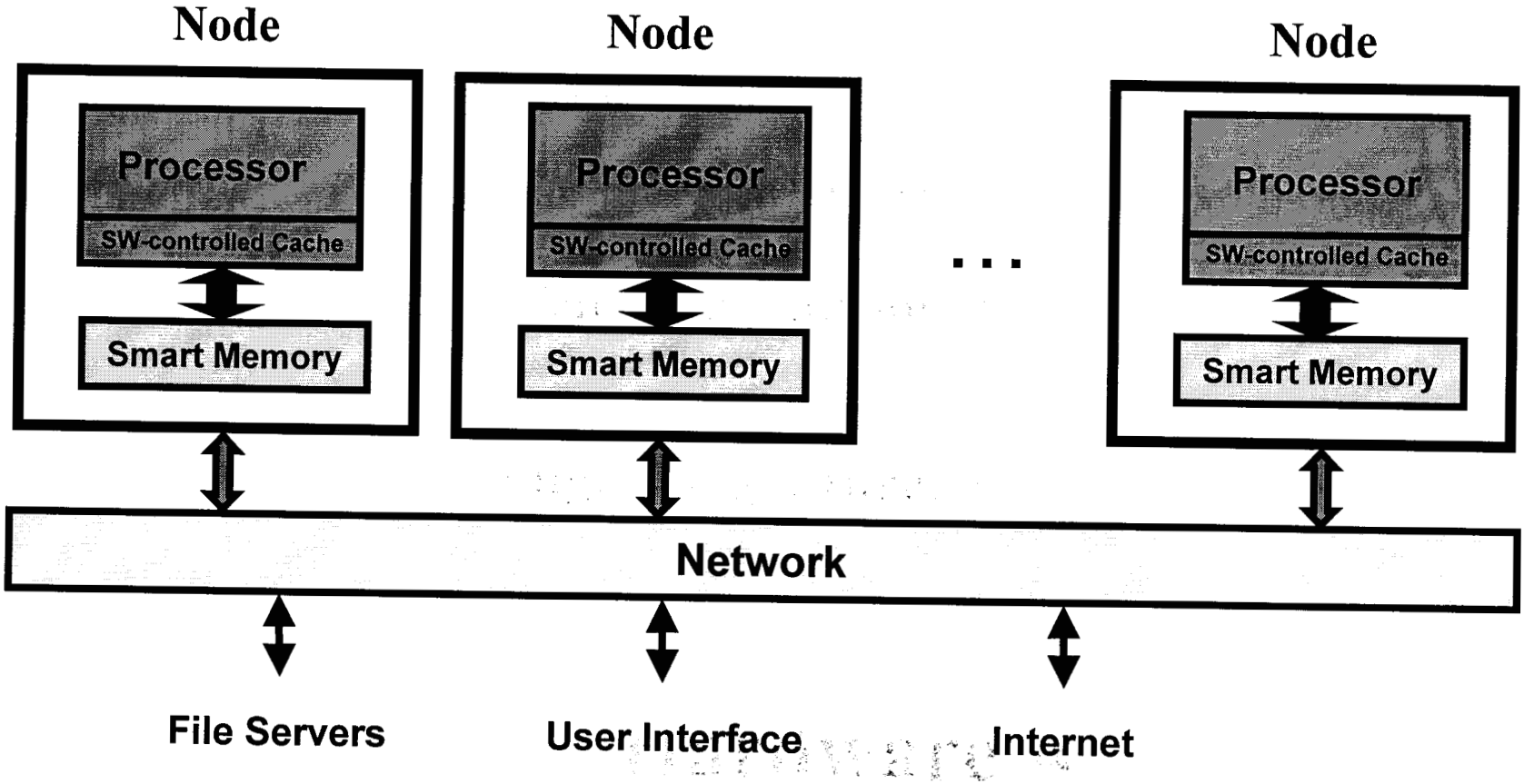
Partners:

- Caltech/JPL
- University of Notre Dame
- Stanford University

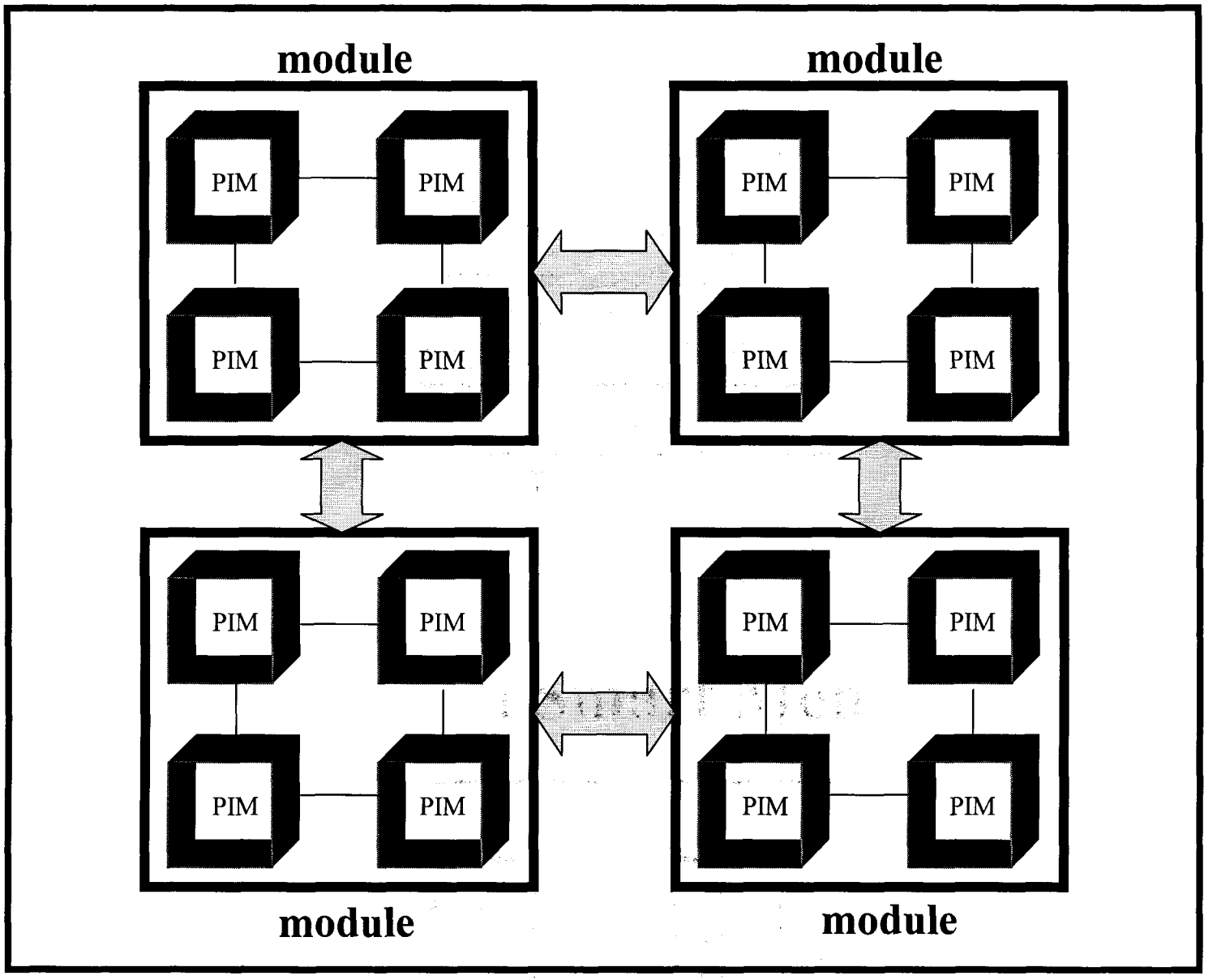
Cascade: Key Elements

- Hierarchical architecture: two levels of processing elements
- Shared address space
- Uniform (UMA) as well as locality-preserving (NUMA) addressing modes
- Smart memory with lightweight threads
- Hybrid programming/execution paradigm
- Fine grain synchronization
- Recovery "on-the-fly"

Cascade Global Hardware Architecture

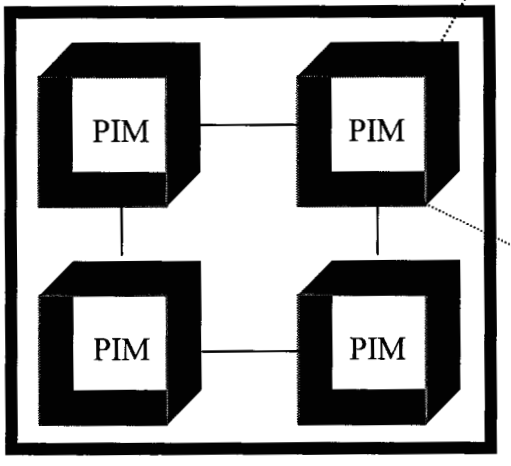


PIM-Based Smart Memory

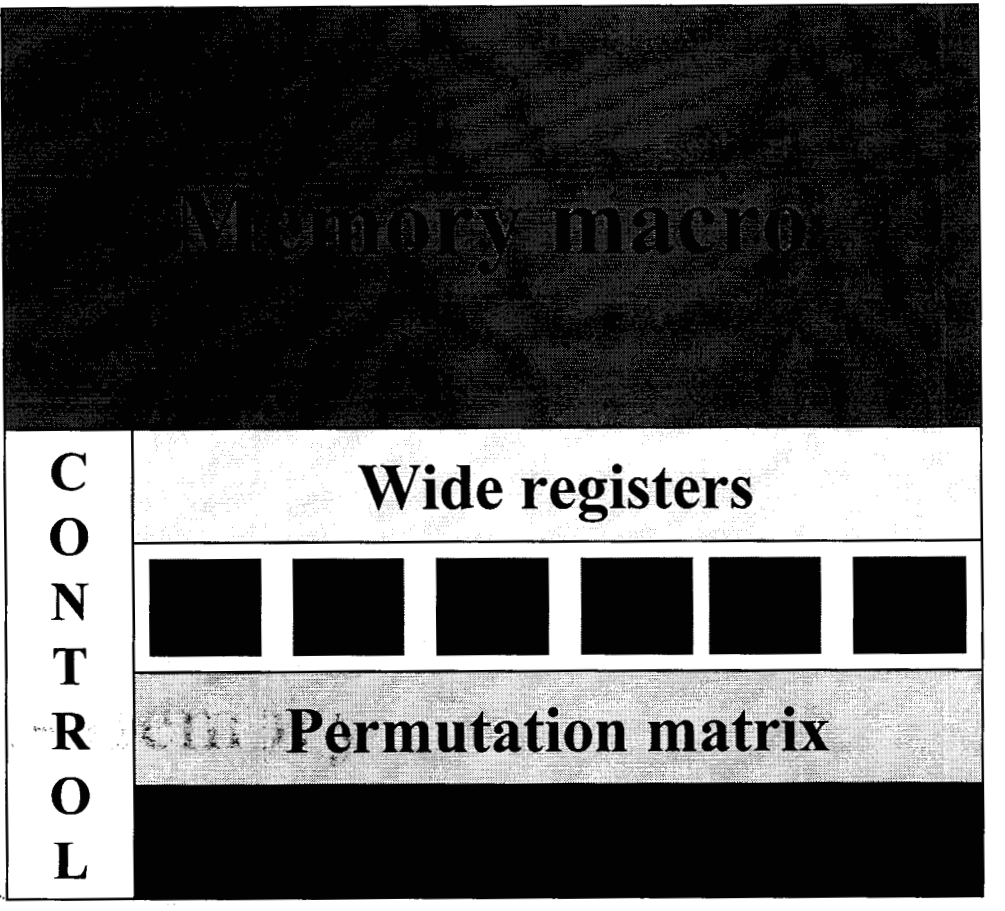


Processor-in-Memory

- *Integration of CMOS processor logic/DRAM memory*
- *Replication of PIM nodes across module*
- *Huge improvement of on-chip bandwidth*
- *Efficient memory operations and wide-word processing*
- *Elimination of data caches*
- *Multithreading support*



PIM module (chip)



PIM node

Cascade 2007 – 1 PetaFlops

- 16K nodes
- Multithreaded Processor
  - 8 GHz clock rate
  - 8-way issue from SMT vectors
  - 64 Gflops peak
- PIM array
  - 8 chips
  - 1 Gigabyte per module
  - 16 GigaFlops peak per chip

Address Translation

- **The address space of an application may contain three different kinds of "segments": (1) globally hashed, (2) locally hashed, and (3) non-hashed**
- **Segments consist of a sequence of virtual *locales*, each of which containing a set of locally translated pages**
- ***Global hashing:* consecutive blocks spread "randomly" across the whole address space**
- **Some of memory can be *locally hashed*, with consecutive blocks spread "randomly" across one locale**
- **Some of memory can be *non-hashed*, with consecutive blocks located within a single memory chip**

Lightweight Threads

- Lightweight threads (LWT) in the memory exploit spatial locality by migrating to the data they refer to
- PIM technology supports LWTs effectively
- LWT are spawned by sending *parcels* to memory
  - Spawning and migration overheads must be minimized
  - In-memory operations are specially supported
- The compiler maps the temporally local loops to Heavyweight threads (HWTs), executed on the node processors, and the others to LWTs

Programming Model Issues

- **Hybrid UMA/NUMA Scheme**
  - **Initial Step based on UMA – *Base Parallel Model***
  - **Tuning via Locality Exploitation - *Extended Parallel Model***
- **Standard languages with simple extensions for parallelism**
- **Directives provide access to advanced features**
- **Tools help bridge the gap to low-level machine model**
- **Execution model supports legacy programs**

Base Parallel Model

- Unbounded lightweight threads
  - Explicit thread creation
  - Special constructs such as "doall" for data parallelism
- Flat shared memory
- Explicit synchronization
- Weak memory semantics: communication must be protected via synchronization

The Base Parallel Model targets high programmer productivity by ignoring machine idiosyncrasies

Base Model Performance Issues

- Amdahl's Law: for fixed sized problems memory latency matters at large scale
- Data bandwidth limited applications cannot be improved by increasing concurrency
- Instruction-fetch limited applications can not be improved by increasing thread-level concurrency
- Cost/performance competitiveness for legacy codes

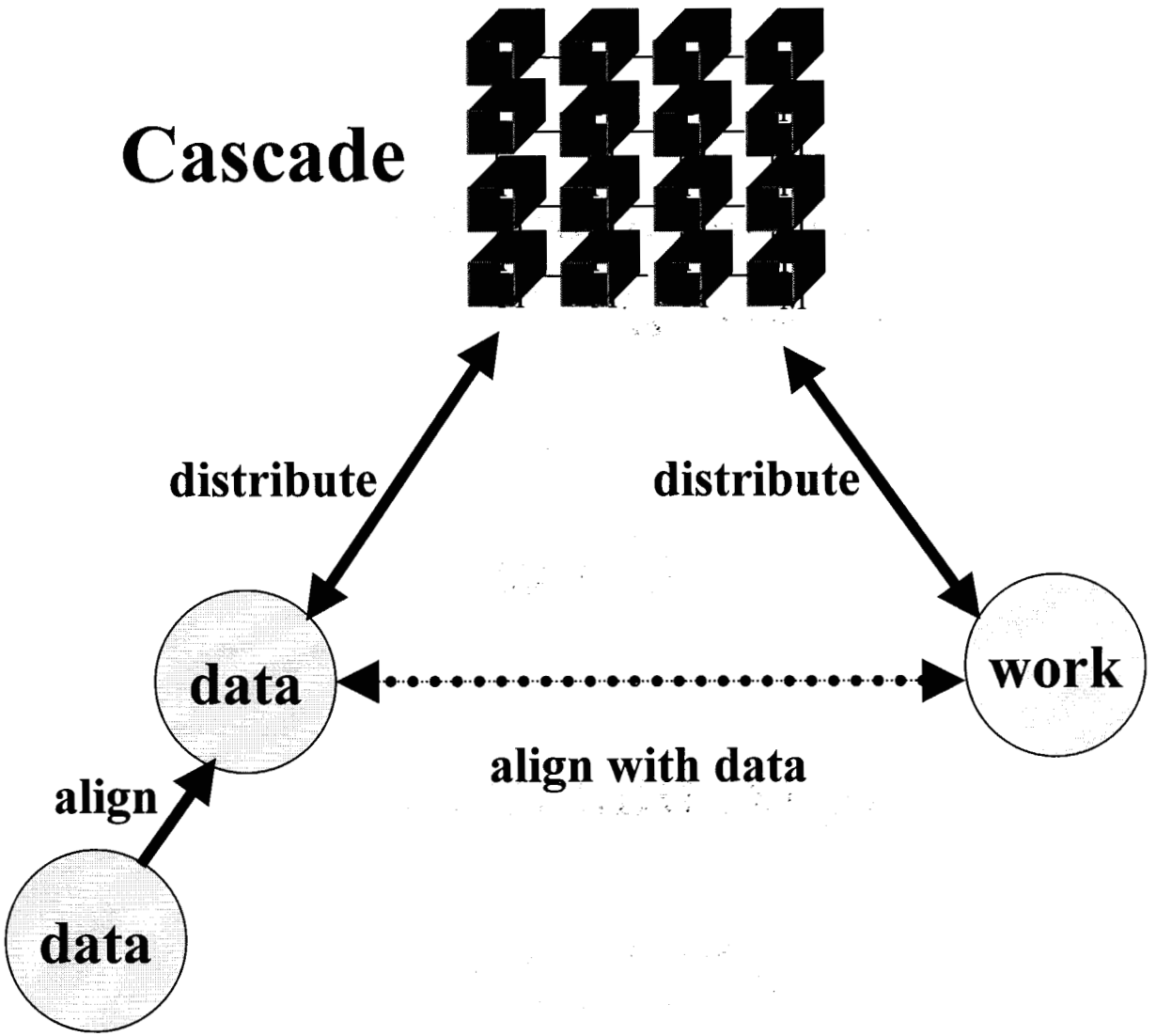
Locality can matter!

Elements of the Extended Programming Model

- ***Abstract architecture specification***
- ***Distribution of data structures to memory/processing elements***
- ***Data alignment***
- ***Data/thread affinity***

Control of these features must be dynamic

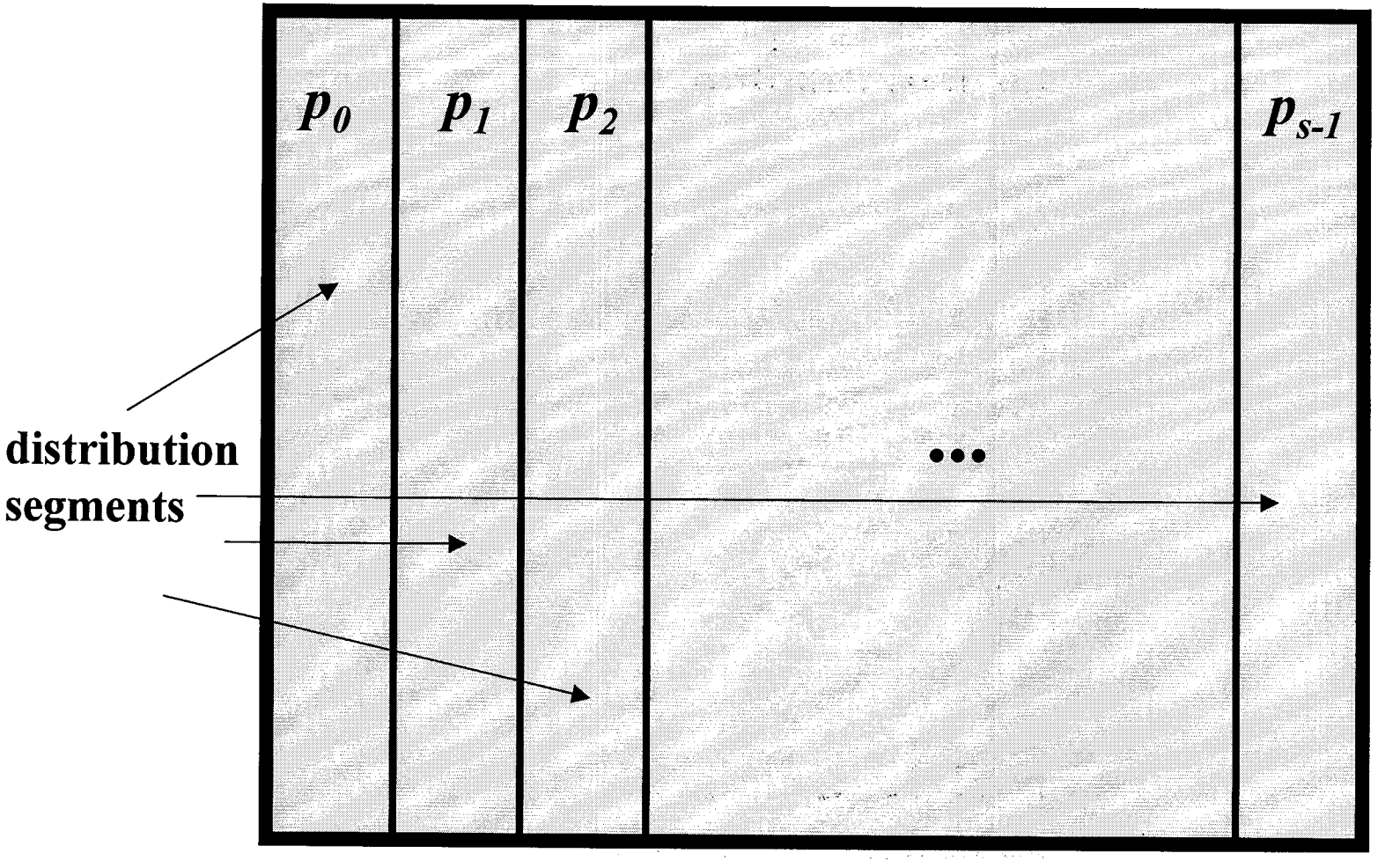
Distribution and Alignment



Collections

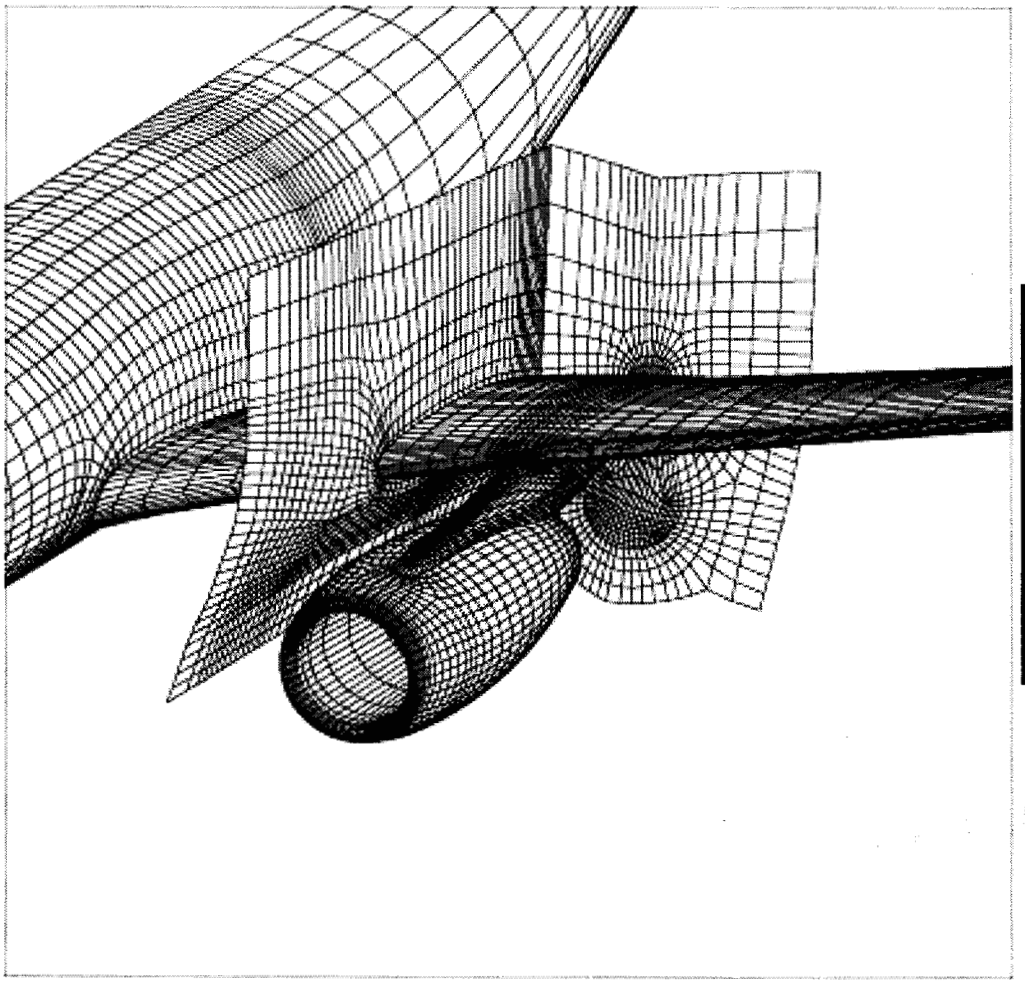
- **Collections (introduced by Sipelstein and Blelloch) are homogeneous or hetereogeneous aggregates, covering a broad range of methods for structuring, naming, and accessing data.**
  - ***(dense) Fortran or C arrays***
  - ***sparse matrices***
  - ***records***
  - ***LISP lists***
  - ***SETL sets***
  - ***mappings***
  - ***graphs and grids***
- **Each collection, C, is associated with a unique index domain, I, which provides a set of unambiguous names for accessing its primitive elements.**
- **A distributed collection is a pair, (C,d), where d:I → U is a distribution of the index domain to a set of memory units.**

Column-block distribution of a 2D-matrix



Regular distributions such as this can be easily handled in the compiler/runtime system

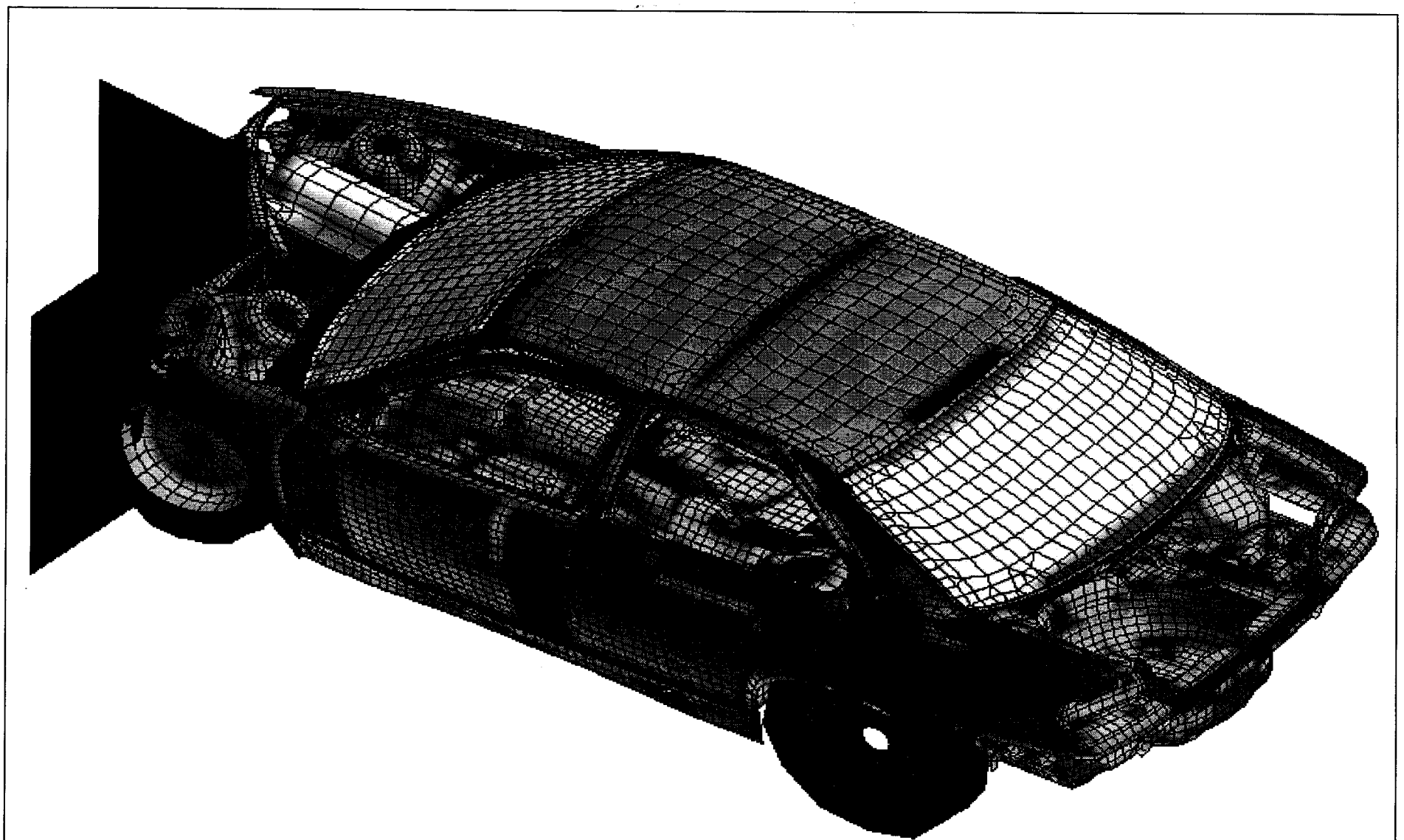
A Multiblock Grid Collection



Source: C.B.Allen, Bristol, UK

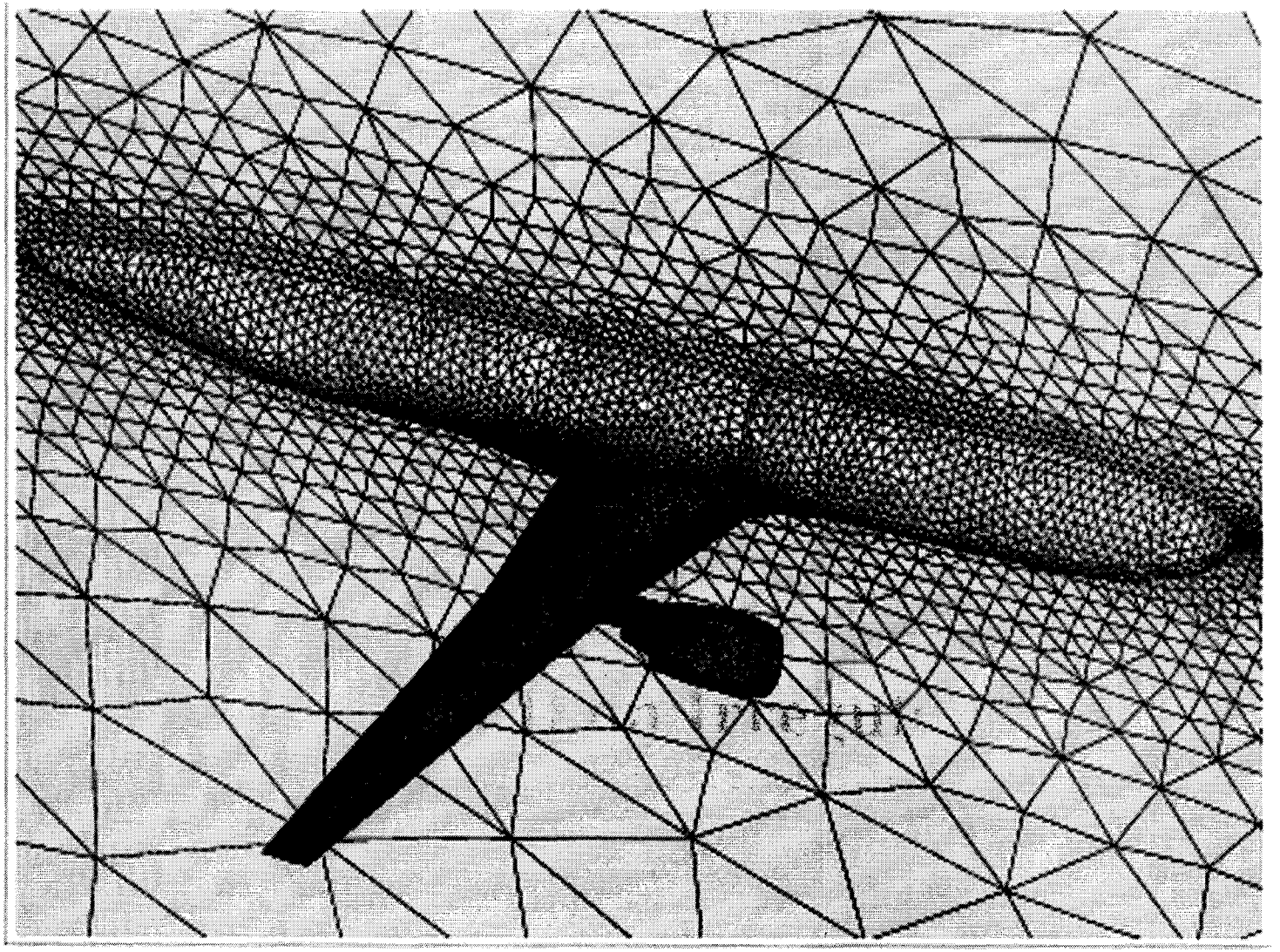
- *define partition of abstract locale set*
- *distribute grids to locale subsets*
- *process grids in parallel across locale subsets*
- *run solvers in parallel on individual locale subsets*

Example: Crash Simulation



Source: Engineering Systems International (ESI)

Example: CFD on Irregular Mesh



Source: Dimitri Mavriplis, ICASE, NASA Langley Research Center

Requirements for Irregular and Dynamic Applications

- **General data structures**
- **General methods for distributing and aligning data**

  (regular distributions may not reflect locality in physical space)

- **General mechanisms for data/thread affinity**

  (allow a dynamic mapping of thread groups to memory segments associated with a data partition)

- **Dynamic manipulation of data distributions, alignments, and affinity must be efficient**

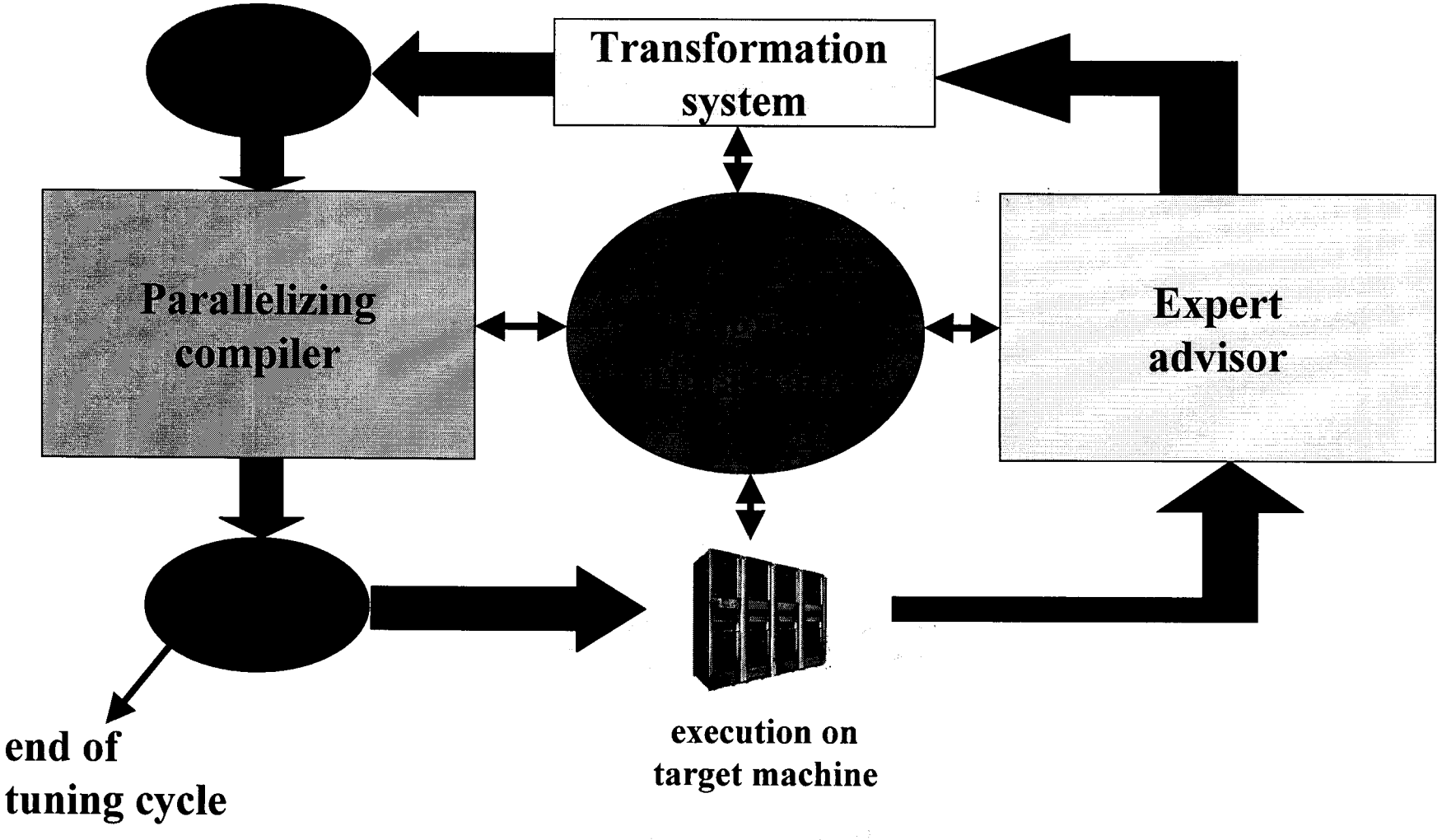
  (apart from adaptive problems such as SAMR dynamic redistributions are even needed for regular problems such as ADI)

Software Infrastructure Components

The user cannot be expected to fully control the system operation at a low level of abstraction, as in today's HPC architectures (e.g., MPI). As a consequence, a set of sophisticated tools for the following functionalities is required:

- ***Automatic distribution***
- ***Directed distribution (a la High Performance Fortran)***
- ***Performance analysis and prediction***
- ***Automatic performance tuning***
- ***High-level debugging***

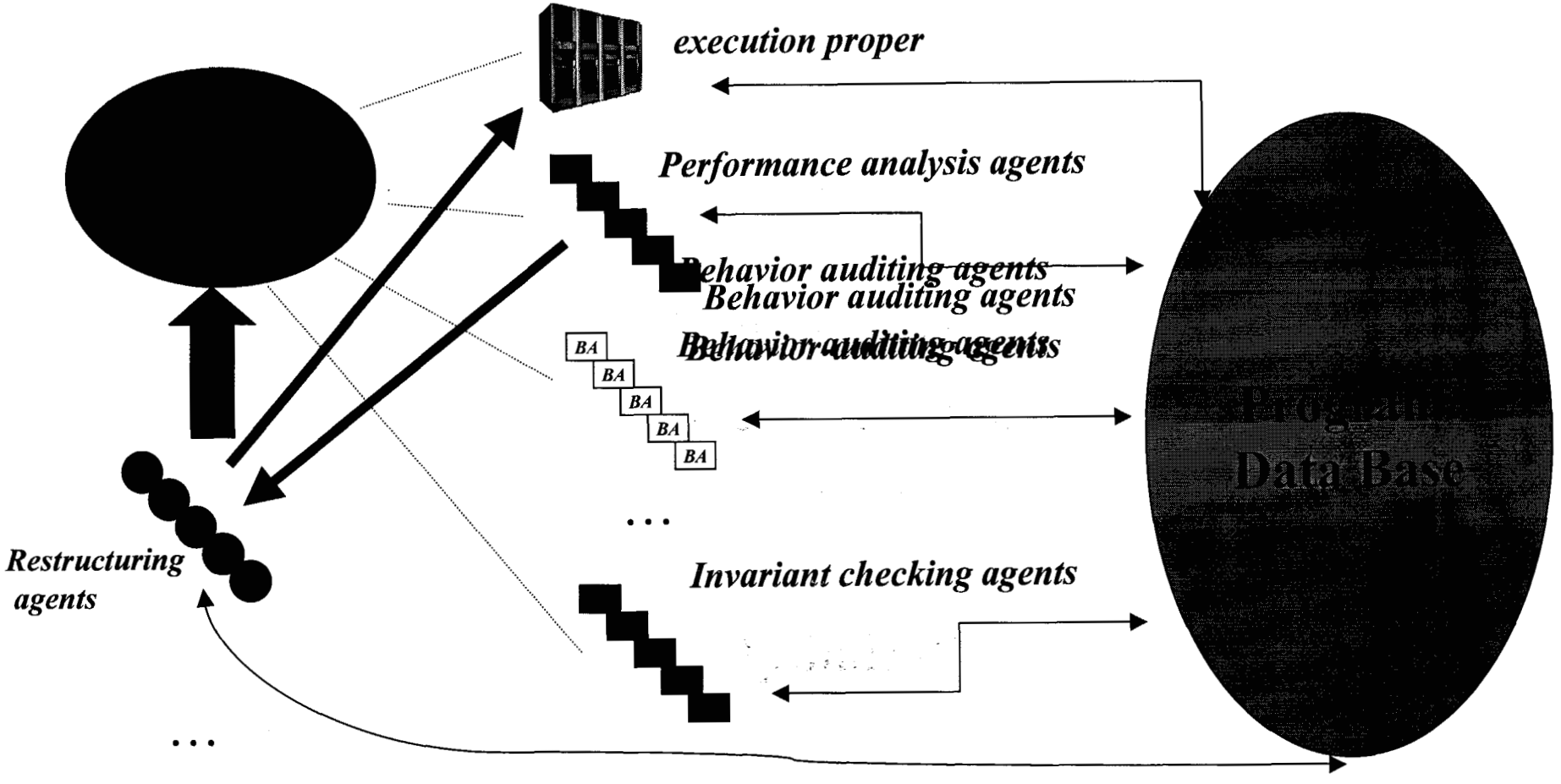
Performance-Guided Offline Tuning



Feedback-Directed Optimization

- Performance-guided off-line tuning is just a point in an optimization continuum
- Other approaches include
  - Runtime code-generation (inspector/executor)
  - On-line optimization in software (Jalapeno,HotSpot)
  - On-line optimization in hardware (trace caches, MTA hotspot strategy)
- Software approaches can use introspection for this purpose

Introspection and Its Use for Optimization and Execution Control



Conclusion

- Cascade is a hierarchical architecture offering a hybrid UMA/NUMA paradigm
- Applications must be parallelized across multiple levels: most of this work must be done by compiler and runtime system, in a user-transparent way
- Leverage of MTA compiler technology and existing NUMA compilation technology is a key to the success of this effort
- Intelligent tools are needed to deal with issues such as performance-guided program restructuring (offline/online)
- Efficient porting of MPI legacy codes will likewise require a sophisticated transformation system with insight into the semantics of the original program (or significant user input)